

Learning Interactive Real-World Simulators

Year: 2023 | Citations: 314 | Authors: Mengjiao Yang, Yilun Du, Kamyar Ghasemipour, Jonathan Tompson, D. Schuurmans

Abstract

Generative models trained on internet data have revolutionized how text, image, and video content can be created. Perhaps the next milestone for generative models is to simulate realistic experience in response to actions taken by humans, robots, and other interactive agents. Applications of a real-world simulator range from controllable content creation in games and movies, to training embodied agents purely in simulation that can be directly deployed in the real world. We explore the possibility of learning a universal simulator (UniSim) of real-world interaction through generative modeling. We first make the important observation that natural datasets available for learning a real-world simulator are often rich along different dimensions (e.g., abundant objects in image data, densely sampled actions in robotics data, and diverse movements in navigation data). With careful orchestration of diverse datasets, each providing a different aspect of the overall experience, we can simulate the visual outcome of both high-level instructions such as "open the drawer" and low-level controls from otherwise static scenes and objects. We use the simulator to train both high-level vision-language policies and low-level reinforcement learning policies, each of which can be deployed in the real world in zero shot after training purely in simulation. We also show that other types of intelligence such as video captioning models can benefit from training with simulated experience, opening up even wider applications. Video demos can be found at <https://universal-simulator.github.io>.